Feature Grouping for Intrusion Detection Based on Mutual Information

Song, Jingping; Zhu, Zhiliang; Price, Christopher

Published in:
Journal of Communications

DOI:
10.12720/jcm.9.12.987-993

Publication date:
2014

Citation for published version (APA):
Song, J., Zhu, Z., & Price, C. (2014). Feature Grouping for Intrusion Detection Based on Mutual Information. *Journal of Communications*, *9*(12), 987-993. https://doi.org/10.12720/jcm.9.12.987-993

Feature Grouping for Intrusion Detection Based on Mutual Information

Jingping Song[1, 2], Zhiliang Zhu[1], and Chris Price[2]
[1]Software College of Northeastern University, Shenyang, 110819, China
[2]Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, United Kingdom
Email: {songjp, zhuzl}@swc.neu.edu.cn; cjp@aber.ac.uk

Abstract **—**Intrusion detection is an important task for network operators in today's Internet. Traditional network intrusion detection systems rely on either specialized signatures of previously seen attacks, or on labeled traffic datasets that are expensive and difficult to re-produce for user-profiling to hunt out network attacks. This paper presents a feature grouping method for the selection of features for intrusion detection. The method is based on mutual information theory and is tested against KDD CUP 99 dataset. It ranks the mutual information between features and uses the fuzzy C means algorithm to compose groups. The largest mutual information between each feature and a class label within a certain group is then selected. The evaluation results show that better classification performance results from such selected features.

Index Terms**—**Mutual information, feature grouping, intrusion detection and feature selection

I. INTRODUCTION

Computer network security has been a very important topic due to the level of attacks and intrusions on networks. One preventive measure that has been deployed on networks to monitor intrusion attacks is the work on Intrusion Detection Systems [1]. An Intrusion Detection System (IDS) is a system that conducts the process of identifying attack behaviour on a network [2]. There are two main detection methods for IDS, anomaly-based [3] or misuse-based [4]. Misuse detection is based on signatures of previously seen attacks that are matched against a stream of audit data looking for evidence of the modelled attacks [5]. The audit data may be obtained from the network, operating systems, or application log files [6]. Misuse-based systems have the advantage of low false positives. Unfortunately, they can only detect those attacks that have been previously specified. In contrast, anomaly-based techniques follow an approach that is complementary to misuse detection [7]. They rely on models, or profiles, of normal users, applications, and network traffic behaviours. Deviations from established models of normal usage are interpreted as attacks. Anomaly detection systems have the advantage that they are able to identify previously unknown attacks.

Classification methods may be used to develop anomaly intrusion detection systems, and machine learning theory can be valuable in this area because of the continued increase of attacks on computer networks. Intrusion detection can be considered as a two class problem or a multiple class problem. A two class problem regards all attack types as anomaly patterns with the rest regarded as a normal pattern. A multiple class problem deals with the classification based on different attacks. For instance, in [8], a method consisting of a combination of discretizers, filters and classifiers is presented. The main goal of that method is to significantly reduce the number of features while maintaining the performance of the classifiers, or even improving it. A mutual information-based feature selection method is reported in [9] that results in detecting intrusions with higher accuracy. Another two feature selection methods have been proposed for intrusion detection systems [10], [11]. Alternative approaches that utilise mutual information theory and feature selection based on this theory have been developed as well [12] and [13].

Feature selection is the process of choosing a subset of the original feature spaces according to discrimination capability in an effect to improve the quality while reducing the dimensionality of data [14]. The number of features extracted from raw network data, which an IDS needs to examine, is usually large even for a small network. Much has been tried in order to increase the detection rate of IDS through proposing new classifiers, but improving the effectiveness of classifiers is not an easy task. However, feature selection can be used to optimize the existing classifiers by removing redundant or irrelevant features. Feature selection is also useful to reduce the computational time and facilitate data understanding. In particular, feature grouping that allows the selection of multiple features by one go is applicable to the dataset with a high dimensionality [15].

Mutual information-based feature selection was initially proposed by in [16] and subsequently modified in 2009 [14] and [15]. This paper proposes a feature selection method by grouping features based on the use of mutual information. The selected features are then employed in the C4.5 classification method [17] for intrusion detection. The performance of the proposed approach is evaluated with respect to different numbers of features and compared with the existing of [14].

II. BACKGROUND

Manuscript received July 28, 2014; revised December 30, 2014.
Corresponding author email: songjp@swc.neu.edu.cn.
doi:10.12720/jcm.9.12.987-993

The purpose of the work in this paper is utilizing feature grouping and mutual information theory to select features and to get better performance evaluation. Network raw datasets usually have large number of features, and feature grouping method can be used to select important features. In this section, mutual information theory is introduced and advantages of feature selection are presented as well.

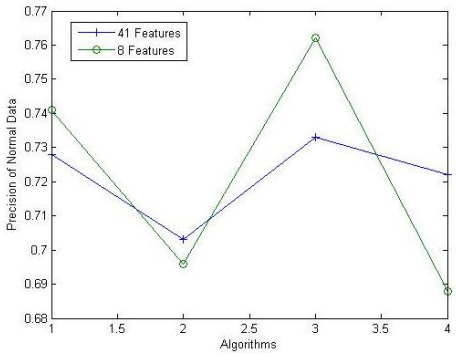

Fig. 1. Precision comparison chart of normal data between all features and selected features.

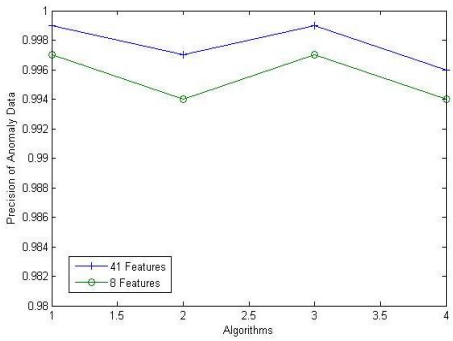

Fig. 2. Precision comparison chart of anomaly data between all features and selected features.

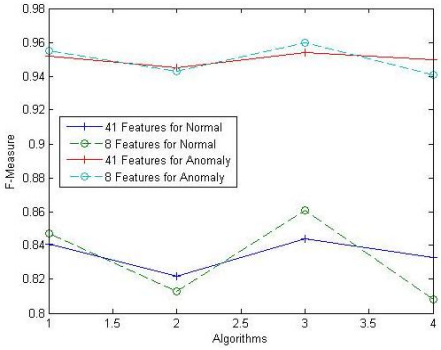

Fig. 3. F-measure comparison chart between all features and selected features.

A. Advantages of Feature Selection

Feature selection is a process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context [18]. In KDD99 dataset, some features may be irrelevant and others may be redundant since the information they add is contained in other features. These extra features can increase computation time for creating classifications, and can have an impact on the accuracy of the classifier built. For this reason, these classification domains seem to be suitable for the application of feature selection methods [19]. These methods are centred in obtaining a subset of features that adequately describe the problem at hand without degrading performance.

To verify that there are irrelevant and redundant features in KDD Cup 99 dataset, Correlation based Feature Selection (CFS) is used to select 8 features by Weka. Two performance measures (precision and F-measure) were calculated which will specifically be discussed in section 4 and four classification methods are used to calculate the two performances. Fig. 1 and Fig. 2 shows the precision comparison between 41 features and 8 features by normal and anomaly types respectively. Similarly, Fig. 3 describes the other performance F-measure.

The three figures show for each classification method, that the two performances are quite close. For two classification algorithm J48 and PART, the performances even get better. Another advantage of selecting features is the running time is shorter than using all features.

B. Definition of Mutual Information

Information theory was initially developed to measure the size of the amount of information in communicating data. And in this theory, entropy is an important measurement for information. It is capable of quantifying the uncertainty of random variables and scaling the amount of information shared by them effectively.

Let X be a random variables with discrete values, its entropy is defined as

$$H(\mathrm{X}) = -\sum_{x \in X} p(\mathrm{x}) \log p(\mathrm{x}) \qquad (1)$$

where H( ) is entropy, and p(x)=Pr(X=x) is the probability density function of X. Note that entropy depends on the probability distribution of the random variable.

Conditional entropy refers to the uncertainty reduction of one variable when the other is known. Assume that variable Y is given, the conditional entropy H(X|Y) of X with respect to Y is

$$H(\mathrm{X} \mid \mathrm{Y}) = -\sum_{y \in Y} \sum_{x \in X} p(\mathrm{x}, \mathrm{y}) \log p(\mathrm{x} \mid \mathrm{y}) \qquad (2)$$

where p(x,y) is the joint probability density function and p(x|y) is the posterior probabilities of X given Y. Similarly, the joint entropy H(X, Y) of X and Y is

$$\begin{aligned} H(\mathrm{X}, \mathrm{Y}) &= H(\mathrm{X}) + H(\mathrm{Y} \mid X) \\ &= H(Y) + H(X \mid Y) \\ &= -\sum_{y \in Y} \sum_{x \in X} p(\mathrm{x}, \mathrm{y}) \log p(\mathrm{x}, \mathrm{y}) \end{aligned} \qquad (3)$$

To quantify how much information is shared by two variables X and Y, a concept termed mutual information I(X; Y) is defined as

$$\begin{aligned} \mathrm{I(X;Y)} &= H(\mathrm{X}) - H(X \mid Y) \\ &= H(Y) - H(Y \mid X) \\ &= -\sum_{y\in Y}\sum_{x\in X} p(\mathrm{x,y}) \log \frac{p(\mathrm{x,y})}{p(\mathrm{x})\,\mathrm{p(y)}} \end{aligned} \tag{4}$$

If X and Y are closely related with each other, I(X; Y) will be very high. Otherwise, I(X; Y)=0 denotes that these two variables are totally unrelated. The mutual information could be applied for evaluating any arbitrary dependency between random variables. In this paper, the mutual information between two variables is calculated and the mutual dependence is measured between them.

III. Proposed Work

Feature selection problem could be described by the context of machine learning. Assume that T=D(F,C) is a training dataset with m instances and n features, where $D=\{o_1,o_2,\ldots o_m\}$and $F=\{f_1,f_2,\ldots f_n\}$ are the sets of instances and features. $C=\{c_1,c_2,\ldots c_k\}$ refers to the set of class labels. For each instance $o_j \in D$, it can be denoted as a value vector of features, i.e., $o_j=(v_{j1},v_{j2},\ldots,v_{jn})$, v_{ji} is the value of o_j corresponding to the feature f_i.

Given a training dataset T=D(F,C), learning algorithms for classification is to induce a hypothesis h: $F_i \rightarrow C$ from T, where F_i is the value domain of $f_i \in F$. Since the limited number of instances in D, there is a classification error $\varepsilon_{F(h)}=|\{(o,c) \in F \ |h(o)\neq c\}|/m$ for each classifier, where h(o) is the predicted class label of o by the hypothesis h.

Feature selection can change F, and result in the changing of $\varepsilon_{F(h)}$. Battiti's work is based on mutual information to select features as follows,

$$\mathrm{I}(f_i;C) - \beta\sum_{f_s\in S} I(f_i;f_s) \tag{5}$$

where f_s is denoted as selected features, and S is represented as the set of selected features. Formula 5 can be used to select the next feature. β is a parameter and determined empirically and Battiti has proposed a value between 0.5 and 1 for β. This algorithm indicates that feature selection should consider not only the mutual information between each feature and class label but also the mutual information between each feature and selected features.

If there are n features in the dataset and f_i is the feature i, then $M_i(f_i;F)$ denotes the mutual information between f_i and all the other features. And it shows in formula 6.

$$M_i(f_i;F)=\sum_{\substack{j=1\\ j\neq i}}^{n} \mathrm{I}(f_i;f_j) \tag{6}$$

When i=1,2,3,…,n, $SUM_{MI}=[\ M_i(f_i;F)\ ]$ denotes the vector set of C.

Feature selection can be improved on through Feature Grouping based on Mutual Information (FGMI) as follows.

Input: A training dataset T=D(F,C).

Output: Selected features S.

(1) Initialize relative parameters: F←'initial set of all features', C ←'class labels', S=∅.

(2) For each feature f_i, calculate the mutual information between f_i and all the other features in F, then sum the results together and it can be calculated by formula 6, and finally get a vector SUM_{MI}.

(3) Rank the SUM_{MI} by Fuzzy C Means algorithm and get G groups.

(4) For each group g in G, calculate mutual information between each feature and class label in C, and then find the maximum value M_g in each group.

(5) Select feature f_s which has the Mg in each group, and put f_s into S, S←'f_s'.

Output the set containing the selected features: S.

From the algorithm shown above, it can be seen that the number of features selected by this algorithm depends on the number of groups. The mutual information between each two features is calculated and Fuzzy C-Means algorithm is used to compose the groups. Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition and unsupervised classification.

At first, the proposed algorithm calculates the mutual information between each feature and all the other features and adding them together, denoted as SUM_{MI}. Then, it ranked the SUM_{MI} by Fuzzy C Means algorithm to get G groups. Moreover, in each group, the algorithm compute mutual information between each feature and class label and get the maximum one. At last, select the feature which has the maximum value.

IV. Experimental Evaluation

In this section, implemented system and the results will be shown. The implemented algorithm will be compared to the Dynamic Mutual Information Feature Selection (DMIFS) algorithm proposed by Huawen Liu. The experiment is tested on KDD99 dataset, which is widely used in IDS domain.

A. KDD 99 Dataset

The KDD 99 dataset was used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD'99 dataset, the fifth International Conference on Knowledge Discovery and Data Mining [20].The competition task was to build a network intrusion detector. This data set is built based on the data captured in DARPA'98 IDS evaluation program. DARPA'98 is about 4 gigabytes of compressed binary tcpdump data of 7 weeks of network traffic, which can be processed into

about 5 million connection records, each with about 100 bytes. The two weeks of test data have around 2 million connection records [21]. The dataset used in this work is a smaller subset, called 10 percent dataset, which contains 494021 instances and it was already used as the training dataset. For the test dataset, the original KDD Cup 99 dataset is used, which is containing 311029 patterns.

A connection is a TCP data packet sequence from start to end in a certain time and data from source IP address to destination IP address in predefined protocol such as TCP or UDP. Each connection is labelled as either normal or attack. The attack type is divided into four categories of 39 types of attacks. The training and test dataset percentages for the four attack categories are shown in Table I. Only 22 types of attacks are in the training dataset, and the other 17 unknown types only occur in the test dataset [22]. It is important to note that the test data is not from the same probability distribution as the training data, and it includes specific attack types not in the training data which make the task more realistic.

TABLE I: PERCENTAGES OF NORMAL CONNECTIONS AND DIFFERENT KINDS OF ATTACKS IN KDD CUP 99

Categories	10% Training dataset (%)	Test dataset
Normal	19.69	19.48
Dos	79.24	73.90
Probe	0.83	1.34
R2L	0.23	5.21
U2R	0.01	0.07

When the KDD Cup 99 dataset is classified, the data can be considered as a binary case or a multiple class case. The binary case regards all attack types as anomaly patterns and the other class is a normal pattern. A multiple class case deals with the classification based on different attacks. In this work, the KDD Cup 99 dataset is treated as a binary case, with the two patterns normal and anomaly data.

B. Performance Evaluation by Different Groups

As described in the section above, the number of features obtained from the algorithm depends on the number of groups. Fuzzy C Means algorithm is used to divide the ranked vector SUM_{MI}. From previous work in this area, the selected features between 8 and 14 could achieve better performance, and the performance evaluations are as follows.

C4.5 algorithm is used to classify the dataset. C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan and it is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. C4.5 uses the concept of information gain to make a tree of classificatory decisions with respect to a previously chosen target classification. The information gain can be described as the effective decrease in entropy resulting from making a choice as to which attribute to use and at what level.

The classification performances are usually denoted by six measures. These six measures are calculated by True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). These measures will be compared by different number of selected features.

True positive rate (TPR): TP/ (TP+FN), also known as detection rate (DR) or sensitivity or recall. Fig. 4 shows the TPR comparison by different number of features.

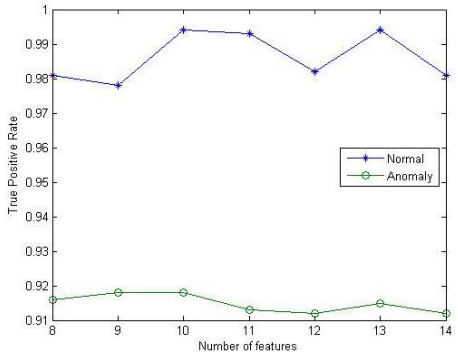

Fig. 4. True positive rate comparison chart by different number of selected features.

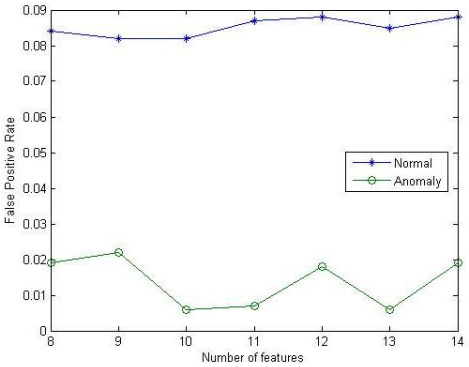

Fig. 5. False positive rate comparison chart by different number of selected features.

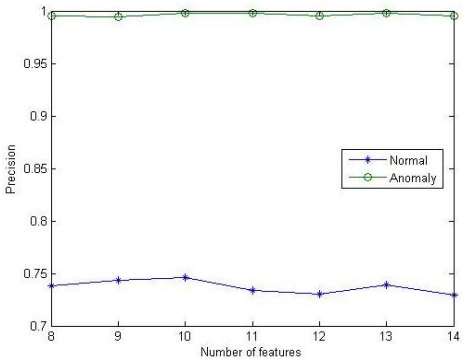

Fig. 6. Precision comparison chart by different number of selected features.

False positive rate (FPR): FP/(TN+FP) also known as the false alarm rate. Fig. 5 describes the FPR comparison by different number of features.

Precision (P): TP/(TP+FP) is defined as the proportion of the true positives against all the positive results. Fig. 6

illustrates the precision comparison by different number of features.

Total Accuracy (TA): (TP+TN)/(TP+TN+FP+FN) is the proportion of true results (both true positives and true negatives) in the population. Recall (R): TP/(TP+FN) is defined as percentage of positive labeled instances that were predicted as positive. F-measure: 2PR/(P+R) is the harmonic mean of precision and recall. The value of Recall is equal to True Positive Rate, and the Recall comparison chart will not be shown. Total Accuracy and F-Measure comparison chart by different number of selected features are shown in Fig. 7 and Fig. 8 respectively.

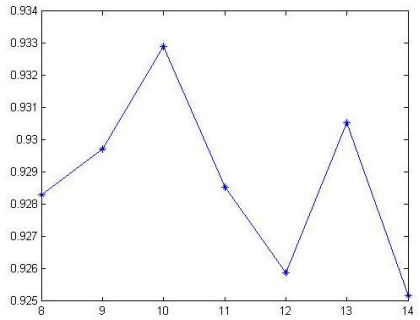

Fig. 7. Total Accuracy comparison chart by different number of selected features.

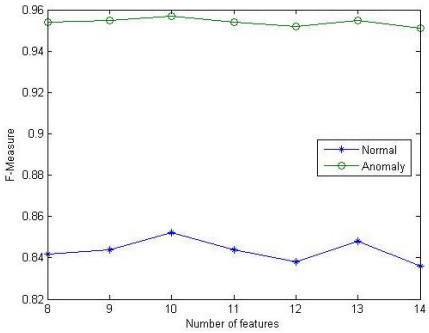

Fig. 8. F-measure comparison chart by different number of selected features.

From the comparison of the measures by different number of selected features, it can be seen that 10 selected features could get the highest TPR, Precision, TA and F-measure. It means 10 selected features could achieve best performance. And 13 selected features could achieve the second best performance.

C. Experiment Results

The experiments were conducted by using KDD 99 dataset and performed on a Windows machine having configuration and Intel (R) Core (TM) i5-2400 CPU@ 3.10GHz, 3.10 GHz, 4GB of RAM, the operating system is Microsoft Windows 7 Professional. And open source machine learning framework Weka 3.5.0 is used to classify the dataset. This tool is used for performance comparison of the proposed algorithm with other classification algorithms. Table II shows the comparison between DMIFS and FGMI. The first row is shown that C4.5 with all 41 features in the dataset. The second row represented DMIFS algorithm proposed by Huawen. 13 features is used by DMIFS and the performance is shown in row 2. The last two rows describe the results of the proposed algorithm FGMI. 13 features and 10 features are used to test by C4.5 respectively. And it is shown from the results that the proposed algorithm could improve the performance of all the measures.

Another 3 algorithms were used to compare beside C4.5, and Table III shows the comparisons by the 3 different classification algorithms. The comparisons are between 41 features and 10 features which are got from the proposed algorithm. The results show that the proposed algorithm could achieve better performance, especially on F-Measure.

One of the advantages of the feature selection method using on KDD 99 dataset is saving computation time. More features means more computation time. Fig. 9 shows the time taken to build model of C4.5 algorithm by different number of features.

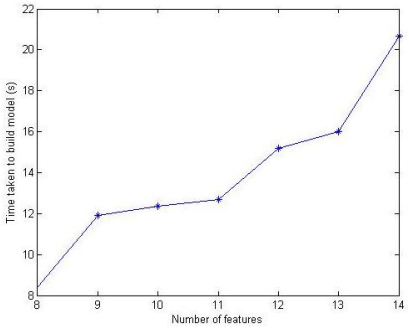

Fig. 9. Time taken to build model comparison chart by different number of features

TABLE II: COMPARISON RESULTS BETWEEN DMIFS AND FGMI.

Algorithm	TP Rate	FP Rate	Precision	F-Measure	Class
C4.5	0.994	0.09	0.728	0.841	Normal
	0.91	0.006	0.999	0.952	Anomaly
C4.5 with DMIFS (13 Features)	0.993	0.086	0.736	0.846	Normal
	0.914	0.007	0.998	0.954	Anomaly
C4.5 with FGMI (13 Features)	0.994	0.085	0.739	0.848	Normal
	0.915	0.006	0.998	0.955	Anomaly
C4.5 with FGMI (10 Features)	0.994	0.082	0.746	0.852	Normal
	0.918	0.006	0.998	0.957	Anomaly

TABLE III: COMPARISONS RESULTS BY DIFFERENT CLASSIFICATION ALGORITHMS.

Algorithm	TP Rate	FP Rate	Precision	F-Measure	Class
PART	0.994	0.087	0.733	0.844	Normal
	0.913	0.006	0.999	0.954	Anomaly
PART	0.982	0.076	0.757	0.855	Normal
(10 Features)	0.924	0.018	0.995	0.958	Anomaly
Bayes	0.976	0.1	0.702	0.817	Normal
	0.9	0.024	0.994	0.944	Anomaly
Bayes	0.979	0.1	0.702	0.818	Normal
(10 features)	0.9	0.021	0.994	0.945	Anomaly
JRip	0.994	0.087	0.734	0.845	Normal
	0.913	0.006	0.998	0.954	Anomaly
JRip	0.982	0.086	0.733	0.84	Normal
(10 features)	0.914	0.018	0.995	0.953	Anomaly

V. CONCLUSION

This paper has presented a feature grouping method based on mutual information. And it specifically proposed how to compose the group by mutual information calculated by each two features, how to get the number of groups and how to rank the features in each group. First of all, the mutual information between one feature and all the other features are calculated to represent the relationship among all the features. Moreover, the proposed algorithm takes advantage of fuzzy C-means algorithm to compose groups. Finally, the mutual information between a feature and class labels are used to select one feature in one group. Experiment results on KDD 99 dataset indicate that the proposed approach generally outperforms DMIFS algorithm. Furthermore, the comparison between 10 features and 41 features by different classification algorithms reveals the performance indicators are improved.

Whilst promising, the presented work opens up an avenue for further investigation. For instance, the mutual information between features and class labels can be used to design new algorithm. And other clustering or classification algorithms can be applied to compose groups. Moreover, more than one feature could be selected in a certain group. In future work, the proposed algorithm will be tested on other datasets and look for more effective measures or methods than mutual information theory.

REFERENCES

[1] S. Shan and V. Karthik, "An approach for automatic selection of relevance features in intrusion detection systems," in *Proc. International Conference on Security and Management*, 2001, pp. 215-219.

[2] S. T. Sobh, "Anomaly detection based on hybrid artificial immune principles," *Information Management & Computer Security*, vol. 21, no. 14, pp. 1-25, 2013.

[3] M. Mehdi, S. Zair, A. Anou, and M. Bensebti, "A bayesian networks in intrusion detection systems," *Journal of Computer Science*, vol. 3, no. 5, pp. 259-265, 2007.

[4] R. M. Rimiru, T. Guanzheng, and S. N. Njuki, "Towards automated intrusion response: A PAMP - based approach," *International Journal of Artificial Intelligence and Expert Systems*, vol. 2, no. 2, pp. 23-35, 2011.

[5] P. Casas, J. Mazel, and P. Owezarski, "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge," *Computer Communications*, vol. 37, no. 7, pp. 772–783, 2012.

[6] I. Onut and A. A. Ghorbani, "A feature classification scheme for network intrusion detection," *International Journal of Network Security*, vol. 5, no. 1, pp. 1-15, 2007.

[7] X. Xu, "Sequential anomaly detection based on temporal-difference learning: Principles, models and case studies," *Applied Soft Computing*, vol. 10, pp. 859–867, 2010.

[8] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset," *Expert Systems with Applications*, vol. 38, pp. 5947–5957, 2011.

[9] A. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835-850, 2005.

[10] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," *Journal of Computers and Security*, vol. 24, no. 4, pp. 295–307, 2005.

[11] S. Mukkamala and A. H. Sung, "Feature ranking and selection for intrusion detection systems using support vector machines," in *Proc. International Conference on Information and Knowledge Engineering*, 2002, pp.503–509.

[12] S. Y. Lee, Y. T. Park, D. Auriol, and J. Brian, "A novel feature selection method based on normalized mutual information," *Applied Intelligence*, vol. 37, no. 1, pp. 100-120, 2012.

[13] F. Benjamin, K. Ammar, H. C. Nabil, C. Chieh, and P. Greg, "Feature selection based on mutual information for human activity recognition," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 1729-1732.

[14] H. W. Liu, J. G. Suna, L.Liu, and H. J. Zhang, "Feature selection with dynamic mutual information," *Pattern Recognition*, vol. 42, 1330-1339, 2009.

[15] F. Amiri, M. M. R. Yousefi, C. Lucas, A. Shakery, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1184–1199, 2011.

[16] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, pp. 537-550, 1994.

[17] A. P. Muniyandia, R. Rajeswarib, and R. Rajaramc, "Network anomaly detection by cascading K-Means clustering and C4.5 decision tree algorithm," in *Proc. International Conference on Communication Technology and System Design*, 2012, pp. 174-182.

[18] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," *Journal of Computers & Security*, vol. 24, no. 4, pp. 295–307, 2005.
[19] S. W. Lin, K. C. Ying, C. Y. Lee, and Z. J. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection," *Applied Soft Computing*, vol. 12, pp. 3285–3290, 2012.
[20] KDD'99 Dataset. University of California, Irvine. [Online]. Available: http:// kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
[21] H. G. Kayacik, Z. Heywood, A. N., and M. I. Heywood, "Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets," in *Proc. Third Annual Conference on Privacy, Security and Trust*, 2005.
[22] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. Second IEEE Symposium on Computational Intelligence for Security and Defence Applications*, 2009.

Jing-ping Song is a lecturer of Software College in Northeastern University in China. He received his master's degree in 2007 from Northeastern University. At the moment, he is working on two PhD subjects. One is about Network Security in Computer Science Department Aberystwyth University UK. The other one is Chaotic Communications for Northeastern University. His research interests are chaos-based digital communication, network security and machine learning. He has published 15 papers and 1 book as co-author.

Zhi-liang Zhu received his PhD degree in computer science from Northeastern University in 2002. His main research interests include information integrate, complexity software system, network coding and communication security, chaos-based digital communications, applications of complex-network theories, and cryptography. By far, he has authored and co-authored over 130 international journal papers and 100 conference papers. Additionally, he published 5 books. He is also the recipient of 9 academic awards at the national, ministerial and provincial level.

Prof. Zhu has served in different capacities in many international journals and conferences. He is a senior member of quite a few professional academic committees, for instance, of the Chinese Institute of Electronics, the China Institute of Communication and the Teaching Guiding Committee for Software Engineering under the Ministry of Education. Moreover, he has led the national natural science fund project, the doctoral program foundation of institution of higher education of china as well as other more than 8 research subjects. For academic exchanges, he has visited to many countries and regions as well as been a visit scholar in the Kansai University, Japan and the San Jose State University, United States.

Chris Price is a Professor in the Department of Computer Science at Aberystwyth University. He is a Fellow of the British Computer Society. He has worked in model-based and qualitative reasoning for the last 25 years, publishing regularly in this field, and iran the EC-funded European Network of Excellence in model-based and qualitative reasoning. His current research interests include model-based diagnosis of unmanned aerial vehicles, efficient ways of building mobile apps, and effective use of knowledge in smart cities.